Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
S1	6413	data adj (mine mining)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 17:24
S2	2854	S1 and (train\$4 learn\$4)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 12:36
<b>S</b> 3	16	S2 and mining adj structur\$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 12:38
<b>S4</b>	74	S2 and bucket\$4	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 12:42
<b>S5</b>	0	S2 and bucket\$4 same (sex male\$2 female\$2) same (salar\$6)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 12:42
<b>S</b> 6	4	S1 same (sex male\$2 female\$2) same (salar\$6)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 12:42
<b>S</b> 7	446	data adj (mine mining) same train\$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 12:44
S8	219	data adj (mine mining) same model\$2 same train\$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 12:44
<b>S</b> 9	228	data adj (mine mining) same model\$3 same train\$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 13:55
S10	7	data adj (mine mining) same model\$3 same train\$3 same bin\$2	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 14:03

C1.		data adi /mina minina) assas anadalah assas basinah adi 11 da	HC DCDHO	OB	ON	2006/04/12 17:21
S11	40	data adj (mine mining) same model\$3 same train\$3 and bin\$2 and classif\$6	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 17:24
S12	7	"6301579" and drill\$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 14:07
S13	1	"6301579".pn. and drill\$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 14:16
S14	1	"6301579".pn. and train\$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 17:20
S15	1	"6301579".pn. and induc\$4	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 17:55
S17	0	"6301579".pn. and (application adj program\$6 api)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 17:21
S18	22	data adj (mine mining) same (train\$4 learn\$4) same (api applicat\$6 adj program\$6 adj interfac\$4)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 19:33
S19	1	"6301579".pn. and "1701"	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 18:59
S20	1	"6865573".pn. and database\$2 with training	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 18:30
S23		"624278".apn. and connect\$5 near3 data	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 19:29
S24	1	"6301579".pn. and indicat\$4	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 19:08

	,	Year				
S25	1	"6301579".pn. and drill\$through	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 19:08
S26	, 1	"624278".apn. and first\$class	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 19:29
S27	0	data adj (mine mining) same (train\$4 learn\$4) same first\$class	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 19:34
S28	2	database with model\$4 with first\$class	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 19:35
S29	30	database same model\$4 same first\$class	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 19:36
S30	32	database same persisten\$4 same first\$class	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 19:38
S31	81	database same first\$class	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 19:38
S32	49	(database same first\$class) not S30	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/12 19:38

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L7	7742	707/104.1.ccls. or 707/102.ccls. or 707/103Y.ccls.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/13 17:11
L8	443	17 and (data adj (mine mining))	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/13 17:11
L9	132	I8 and train\$4	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2006/01/13 17:11



Subscribe (Full Service) Register (Limited Service, Free) Login

Search: 

The ACM Digital Library 
The Guide

+"data mining" +"training" +"bin"

SEARCH



Feedback Report a problem Satisfaction survey

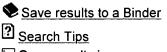
Terms used data mining training bin

Found **120** of **169,166** 

Sort results by Display

results

relevance expanded form



Try an <u>Advanced Search</u>
Try this search in <u>The ACM Guide</u>

next

Open results in a new window

Results 1 - 20 of 120

Result page: 1 2 3 4 5 6 7

Relevance scale

1 Reviewed articles: An internet routing forensics framework for discovering rules of



abnormal BGP events

Jun Li, Dejing Dou, Zhen Wu, Shiwoong Kim, Vikash Agarwal

October 2005 ACM SIGCOMM Computer Communication Review, Volume 35 Issue 5

Publisher: ACM Press

Full text available: pdf(310.62 KB) Additional Information: full citation, abstract, references, index terms

Abnormal BGP events such as attacks, misconfigurations, electricity failures, can cause anomalous or pathological routing behavior at either global level or prefix level, and thus must be detected in their early stages. Instead of using ad hoc methods to analyze BGP data, in this paper we introduce an Internet Routing Forensics framework to systematically process BGP routing data, discover rules of abnormal BGP events, and apply these rules to detect the occurrences of these events. In particula ...

**Keywords**: abnormal BGP events, blackout, data mining, internet worms, routing forensics

2 Bioinformatics: Weave amino acid sequences for protein secondary structure



prediction

Xiaochun Yang, Bin Wang

June 2003 Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery

Publisher: ACM Press

Full text available: pdf(160.75 KB) Additional Information: full citation, abstract, references, index terms

Given a known protein sequence, predicting its secondary structure can help understand its three-dimensional (tertiary) structure, i.e., the folding. In this paper, we present an approach for predicting protein secondary structures. Different from the existing prediction methods, our approach proposes an encoding schema that weaves physiochemical information in encoded vectors and a prediction framework that combines the context information with secondary structure segments. We employed Support ...

**Keywords**: SVM, encoding schema, protein secondary structure, protein structure prediction

Predictor Models in Software Engineering (PROMISE): Nearest neighbor sampling





## for better defect prediction

Gary D. Boetticher

May 2005 ACM SIGSOFT Software Engineering Notes, Proceedings of the 2005 workshop on Predictor models in software engineering PROMISE '05,

Volume 30 Issue 4

**Publisher: ACM Press** 

Full text available: pdf(60.09 KB) Additional Information: full citation, abstract, references, index terms

An important step in building effective predictive models applies one or more sampling techniques. Traditional sampling techniques include random, stratified, systemic, and clustered. The problem with these techniques is that they focus on the class attribute, rather than the non-class attributes. For example, if a test instance's nearest neighbor is from the opposite class of the training set, then it seems doomed to misclassification. To illustrate this problem, this paper conduc ...

Keywords: NASA data repository, decision trees, defect prediction, empirical software engineering, nearest neighbor analysis

Data mining of multidimensional remotely sensed images



Robert F. Cromp, William J. Campbell

December 1993 Proceedings of the second international conference on Information and knowledge management

Publisher: ACM Press

Full text available: pdf(1.39 MB) Additional Information: full citation, references, citings, index terms

Classification and regression: money \*can\* grow on trees



Johannes Gehrke, Wie-Yin Loh, Raghu Ramakrishnan

August 1999 Tutorial notes of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining

Publisher: ACM Press

Full text available: pdf(2.95 MB)

Additional Information: full citation, abstract, references, citings, index terms

With over 800 million pages covering most areas of human endeavor, the World-wide Web is a fertile ground for data mining research to make a difference to the effectiveness of information search. Today, Web surfers access the Web through two dominant interfaces clicking on hyperlinks and searching via keyword queries This process is often tentative and unsatisfactory Better support is needed for expressing one's information need and dealing with a search result in more structured ways than ...

6 Special section on data mining for intrusion detection and threat analysis: ADAM: a



testbed for exploring the use of data mining in intrusion detection

Daniel Barbará, Julia Couto, Sushil Jajodia, Ningning Wu December 2001 ACM SIGMOD Record, Volume 30 Issue 4

Publisher: ACM Press

Full text available: pdf(896.53 KB) Additional Information: full citation, abstract, references, index terms

Intrusion detection systems have traditionally been based on the characterization of an attack and the tracking of the activity on the system to see if it matches that characterization. Recently, new intrusion detection systems based on data mining are making their appearance in the field. This paper describes the design and experiences with the ADAM (Audit Data Analysis and Mining) system, which we use as a testbed to study how useful data mining techniques can be in intrusion detection.

**Keywords**: cyber attacks, data mining, intrusion detection software, pattern analysis

7 Research track: Mining concept-drifting data streams using ensemble classifiers Haixun Wang, Wei Fan, Philip S. Yu, Jiawei Han



August 2003 Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining

**Publisher: ACM Press** 

Full text available: pdf(234.13 KB)

Additional Information: full citation, abstract, references, citings, index terms

Recently, mining data streams with concept drifts for actionable insights has become an important and challenging task for a wide range of applications including credit card fraud protection, target marketing, network intrusion detection, etc. Conventional knowledge discovery tools are facing two challenges, the overwhelming volume of the streaming data, and the concept drifts. In this paper, we propose a general framework for mining concept-drifting data streams using weighted ensemble classifi ...

Keywords: classifier, classifier ensemble, concept drift, data streams

8 Poster papers: Transforming classifier scores into accurate multiclass probability



, <u>estimates</u>

Bianca Zadrozny, Charles Elkan

July 2002 Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining

Publisher: ACM Press

Full text available: pdf(690.25 KB)

Additional Information: full citation, abstract, references, citings, index terms

Class membership probability estimates are important for many applications of data mining in which classification outputs are combined with other sources of information for decision-making, such as example-dependent misclassification costs, the outputs of other classifiers, or domain knowledge. Previous calibration methods apply only to two-class problems. Here, we show how to obtain accurate probability estimates for multiclass problems by combining calibrated binary probability estimates. We a ...

Research track: Eliminating noisy information in Web pages for data mining



Lan Yi, Bing Liu, Xiaoli Li

August 2003 Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining

**Publisher: ACM Press** 

Full text available: pdf(468.87 KB)

Additional Information: full citation, abstract, references, citings, index

A commercial Web page typically contains many information blocks. Apart from the main content blocks, it usually has such blocks as navigation panels, copyright and privacy notices, and advertisements (for business purposes and for easy user access). We call these blocks that are not the main content blocks of the page the noisy blocks. We show that the information contained in these noisy blocks can seriously harm Web data mining. Eliminating these noises is thus of great importance. In this pa ...

**Keywords**: Web mining, noise detection, noise elimination

The true lift model: a novel data mining approach to response modeling in database marketing





Victor S. Y. Lo

December 2002 ACM SIGKDD Explorations Newsletter, Volume 4 Issue 2

Publisher: ACM Press

Full text available: pdf(119.81 KB) Additional Information: full citation, abstract, references

In database marketing, data mining has been used extensively to find the optimal customer targets so as to maximize return on investment. In particular, using marketing campaign data, models are typically developed to identify characteristics of customers who are most likely to respond. While these models are helpful in identifying the likely responders, they may be targeting customers who have decided to take the desirable action or not regardless of whether they receive the campaign contact (e ...

**Keywords**: customer development, customer relationship management, data mining, database marketing, interaction effect, knowledge discovery, predictive modeling, response modeling, treatment effect, true lift, upselling and cross-selling

11 Targeting the right students using data mining



Yiming Ma, Bing Liu, Ching Kian Wong, Philip S. Yu, Shuik Ming Lee

August 2000 Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining

Publisher: ACM Press

Full text available: pdf(108.79 KB) Additional Information: full citation, references, citings, index terms

Keywords: data mining application in education, scoring, target selection

12 Industry track papers: ADMIT: anomaly-based data mining for intrusions



Karlton Sequeira, Mohammed Zaki

July 2002 Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining

Publisher: ACM Press

Full text available: pdf(1.33 MB)

Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>citings</u>, <u>index</u> terms

Security of computer systems is essential to their acceptance and utility. Computer security analysts use intrusion detection systems to assist them in maintaining computer system security. This paper deals with the problem of differentiating between masqueraders and the true user of a computer terminal. Prior efficient solutions are less suited to real time application, often requiring all training data to be labeled, and do not inherently provide an intuitive idea of what the data model means. ...

13 Fast detection of communication patterns in distributed executions



November 1997 Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research

Publisher: IBM Press

Full text available: pdf(4.21 MB) Additional Information: full citation, abstract, references, index terms

Understanding distributed applications is a tedious and difficult task. Visualizations based on process-time diagrams are often used to obtain a better understanding of the execution of the application. The visualization tool we use is Poet, an event tracer developed at the University of Waterloo. However, these diagrams are often very complex and do not provide the user with the desired overview of the application. In our experience, such tools display repeated occurrences of non-trivial commun ...

sensors for detecting abnormal computer usage

14 Research track papers: Selection, combination, and evaluation of effective software



Jude Shavlik, Mark Shavlik

August 2004 Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining KDD '04

Publisher: ACM Press

Full text available: pdf(192.20 KB) Additional Information: full citation, abstract, references, index terms

We present and empirically analyze a machine-learning approach for detecting intrusions on individual computers. Our Winnow-based algorithm continually monitors user and system behavior, recording such properties as the number of bytes transferred over the last 10 seconds, the programs that currently are running, and the load on the CPU. In all, hundreds of measurements are made and analyzed each second. Using this data, our algorithm creates a model that represents each particular computer's ra ...

**Keywords:** Windows 2000, Winnow algorithm, anomaly detection, feature selection, intrusion detection, machine learning, user modeling

15 Research track papers: Data mining in metric space: an empirical analysis of



supervised learning performance criteria

Rich Caruana, Alexandru Niculescu-Mizil

August 2004 Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining KDD '04

Publisher: ACM Press

Full text available: pdf(267.16 KB) Additional Information: full citation, abstract, references, index terms

Many criteria can be used to evaluate the performance of supervised learning. Different criteria are appropriate in different settings, and it is not always clear which criteria to use. A further complication is that learning methods that perform well on one criterion may not perform well on other criteria. For example, SVMs and boosting are designed to optimize accuracy, whereas neural nets typically optimize squared error or cross entropy. We conducted an empirical study using a variety of lea ...

**Keywords**: ROC, cross entropy, lift, metrics, performance evaluation, precision, recall, supervised learning

16 A framework for constructing features and models for intrusion detection systems



Wenke Lee, Salvatore J. Stolfo

November 2000 ACM Transactions on Information and System Security (TISSEC),
Volume 3 Issue 4

**Publisher: ACM Press** 

Full text available: pdf(187.03 KB)

Additional Information: full citation, abstract, references, citings, index terms, review

Intrusion detection (ID) is an important component of infrastructure protection mechanisms. Intrusion detection systems (IDSs) need to be accurate, adaptive, and extensible. Given these requirements and the complexities of today's network environments, we need a more systematic and automated IDS development process rather that the pure knowledge encoding and engineering approaches. This article describes a novel framework, MADAM ID, for Mining Audit Data for Automated Models for Instrusion ...

**Keywords**: data mining, feature construction, intrusion detection

## 17 Research track papers: Towards parameter-free data mining



Eamonn Keogh, Stefano Lonardi, Chotirat Ann Ratanamahatana

August 2004 Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining KDD '04

Publisher: ACM Press

Full text available: Top pdf(770.63 KB) Additional Information: full citation, abstract, references, index terms

Most data mining algorithms require the setting of many input parameters. Two main dangers of working with parameter-laden algorithms are the following. First, incorrect settings may cause an algorithm to fail in finding the true patterns. Second, a perhaps more insidious problem is that the algorithm may report spurious patterns that do not really exist, or greatly overestimate the significance of the reported patterns. This is especially likely when the user fails to understand the role of par ...

**Keywords**: anomaly detection, clustering, parameter-free data mining

18 Learning and making decisions when costs and probabilities are both unknown



Biança Zadrozny, Charles Elkan

August 2001 Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining

Publisher: ACM Press

Full text available: pdf(920.23 KB)

Additional Information: full citation, abstract, references, citings, index terms

In many data mining domains, misclassification costs are different for different examples, in the same way that class membership probabilities are example-dependent. In these domains, both costs and probabilities are unknown for test examples, so both cost estimators and probability estimators must be learned. After discussing how to make optimal decisions given cost and probability estimates, we present decision tree and naive Bayesian learning methods for obtaining well-calibrated probability ...

19 Research track papers: Systematic data selection to mine concept-drifting data





streams Wei Fan

August 2004 Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining KDD '04

**Publisher: ACM Press** 

Full text available: pdf(226.95 KB) Additional Information: full citation, abstract, references, index terms

One major problem of existing methods to mine data streams is that it makes ad hoc choices to combine most recent data with some amount of old data to search the new hypothesis. The assumption is that the additional old data always helps produce a more accurate hypothesis than using the most recent data only. We first criticize this notion and point out that using old data blindly is not better than "gambling"; in other words, it helps increase the accuracy only if we are "lucky." We discuss and ...

**Keywords**: concept-drift, data streams, decision trees

20 Industry/government track paper: Using relational knowledge discovery to prevent



securities fraud

Jennifer Neville, Özgür Şimşek, David Jensen, John Komoroske, Kelly Palmer, Henry Goldberg

August 2005 Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining KDD '05

**Publisher: ACM Press** 

Full text available: pdf(1.22 MB) Additional Information: full citation, abstract, references, index terms

We describe an application of relational knowledge discovery to a key regulatory mission of the National Association of Securities Dealers (NASD). NASD is the world's largest private-sector securities regulator, with responsibility for preventing and discovering misconduct among securities brokers. Our goal was to help focus NASD's limited regulatory resources on the brokers who are most likely to engage in securities violations. Using statistical relational learning algorithms, we developed mod ...

Keywords: fraud detection, relational probability trees, statistical relational learning

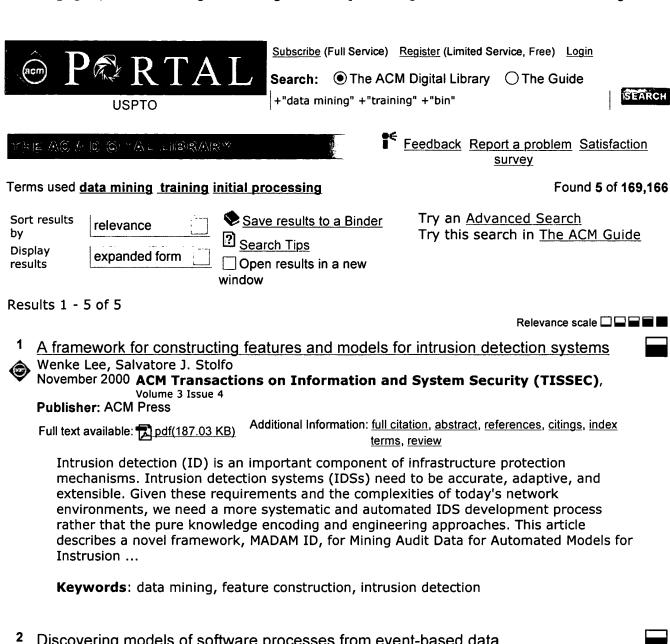
Results 1 - 20 of 120

Result page: 1 2 3 4 5 6 7 next

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2006 ACM, Inc.

Terms of Usage Privacy Policy Code of Ethics Contact Us

Useful downloads: Adobe Acrobat Q QuickTime Windows Media Player Real Player



2 Discovering models of software processes from event-based data

Jonathan E. Cook, Alexander L. Wolf

July 1998 ACM Transactions on Software Engineering and Methodology (TOSEM), Volume 7 Issue 3

Publisher: ACM Press

Additional Information: full citation, abstract, references, citings, index Full text available: pdf(369.76 KB) terms, review

Many software process methods and tools presuppose the existence of a formal model of a process. Unfortunately, developing a formal model for an on-going, complex process can be difficult, costly, and error prone. This presents a practical barrier to the adoption of process technologies, which would be lowered by automated assistance in creating formal models. To this end, we have developed a data analysis technique that we term process discovery. Under this technique, data ...

**Keywords**: Balboa, process discovery, software process, tools

Visualizing geospatial data Theresa Marie Rhyne, Alan MacEachern, Theresa-Marie Rhyne



## August 2004 Proceedings of the conference on SIGGRAPH 2004 course notes GRAPH

Publisher: ACM Press

Full text available: pdf(13.99 MB) Additional Information: full citation, abstract

This course reviews concepts and highlights new directions in GeoVisualization. We review four levels of integrating geospatial data and geographic information systems (GIS) with scientific and information visualization (VIS) methods. These include: • Rudimentary: minimal data sharing between the GIS and Vis systems. Operational: consistency of geospatial data. Functional: transparent communication between the GIS and Vis systems. Merged: one comprehensive toolkit environmentW ...

4 Organizational issues: Behind the help desk: evolution of a knowledge management



system in a large organization

Christine A. Halverson, Thomas Erickson, Mark S. Ackerman

November 2004 Proceedings of the 2004 ACM conference on Computer supported cooperative work

**Publisher: ACM Press** 

Full text available: pdf(413.72 KB) Additional Information: full citation, abstract, references, index terms

This paper examines the way in which a knowledge management system (KMS)-by which we mean the people, processes and software-came into being and evolved in response to a variety of shifting social, technical and organizational pressures. We draw upon data from a two year ethnographic study of a sophisticated help desk to trace the KMS from its initial conception as a "Common Problems" database for help desk personnel, to its current instantiation as a set of Frequently Asked Questions publish ...

Keywords: design approaches, distributed cognition, ethnography, frequently asked questions, help desk

5 The object flow model for data-based simulation



Lois M. L. Delcambre, Lissa F. Pollacia

December 1993 Proceedings of the 25th conference on Winter simulation

Publisher: ACM Press

Full text available: pdf(687.73 KB) Additional Information: full citation, references, citings

Results 1 - 5 of 5

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2006 ACM, Inc. Terms of Usage Privacy Policy Code of Ethics Contact Us

Useful downloads: Adobe Acrobat QuickTime Windows Media Player Real Player